Self-Supervised Learning on Graphs: Contrastive, Generative, or Predictive

Abstract

Deep learning on graphs has recently achieved remarkable success on a variety of tasks, while such success relies heavily on the massive and carefully labeled data. However, precise annotations are generally very expensive and time-consuming. To address this problem, self-supervised learning (SSL) is emerging as a new paradigm for extracting informative knowledge through well-designed pretext tasks without relying on manual labels. In this survey, we extend the concept of SSL, which first emerged in the fields of computer vision and natural language processing, to present a timely and comprehensive review of existing SSL techniques for graph data. Specifically, we divide existing graph SSL methods into three categories: contrastive, generative, and predictive. More importantly, unlike other surveys that only provide a high-level description of published research, we present an additional mathematical summary of existing works in a unified framework. Furthermore, to facilitate methodological development and empirical comparisons, we also summarize the commonly used datasets, evaluation metrics, downstream tasks, open-source implementations, and experimental study of various algorithms. Finally, we discuss the technical challenges and potential future directions for improving graph self-supervised learning. Latest advances in graph SSL are summarized in a GitHub repository https://github.com/LirongWu/awesome-graph-self-supervised-learning.